Technical Bulletin

Part No. 74-0112

DataStage Informix Load

This technical bulletin describes Release 2.1 of the DataStage Informix Load Plug-in, formerly called the DataStage Informix Bulk Load Plug-in. This stage loads data from any DataStage stage into an Informix Warehouse database.

Printing History

First Edition (74-0112) for Release 1.0, August 1997
Second Edition (74-0112) for Release 1.1, March 1999
Third Edition (74-0112) for Release 2.0, July 2000
Updated (74-0112) for Release 2.1, September 2001
Updated (74-0112 for Release 2.1, August 2002
Updated for Release 2.1, November 2002
Updated for Release 2.1, August 2003

How to Order Technical Documents

To order copies of documents, contact your local Ascential subsidiary or distributor, or call our office at (508) 366-3888.

Documentation Team: Marie E. Hedin

Introduction

This technical bulletin describes the following for Release 2.1 of the Informix Load stage, updated for DataStage Release 7.0:

The Informix Load stage enables you to rapidly and efficiently prepare and load streams of tabular data from any DataStage stage (for example, the ODBC stage, the Sequential File stage, and so forth) into tables of the target Informix database for Release 7.2 and later.

Informix Load is a passive stage that uses the Informix DBLOAD utility.

This stage supports one or more stream input links (but no output links). Each input link corresponds to a different bulk loading session within a DataStage job.

Functionality

The Informix Load stage has the following functionality:

- Support for data files that exceed the 2-GB file size limit for 64-bit file systems.
- An option to use the Informix DBLOAD utility for automatic bulk loading. This is useful when you run a DataStage job from an administrative account on the same machine as the Informix server.
- Multiple bulk-loading operations through a single Informix Load stage.
- Before- and after-link external user routines that can use arguments for link properties that you specify. The values for these link properties, such as Table Name, Database Name, Control File Name, and Data File Name, are substituted at run time. For more information about these values, see "Link Properties" on page 6.

  You can use these external routines for before-link operations (for example, initiating customized table cleanup) or after-link operations (for example, initiating custom or remote bulk loading, deleting temporary files, initiating

file transfer and remote processing, recreating table indexes that were dropped before bulk loading, and so forth).

- Support for MetaStage. For more information, see *MetaStage User's Guide.*
- Support for NLS (National Language Support). For more information, see *DataStage NLS Guide*.

The following functionality is not supported:

- Bulk loading multiple tables from a single link (you can do this only by providing a customized control file).
- Generation of fixed-record data file format.
- Multiple data files for a single bulk loading session (you can do this only by using multiple links through the Informix Load stage).
- Automatic bulk loading option works only when you run a DataStage job from an administrative account on the same machine as the Windows NT Informix server. Otherwise, you can also use automatic bulk loading by providing an external after-link routine.

  The bulk loading process does not clean up the table before loading. Use an external before-link routine to start the cleanup of the table.

Terminology

The following table explains the Informix terms used in this document:

Term	Description
before-link and after-link routines	A user-specified external routine or process that performs an action before or after all link-related activities for a specific link, for example, cleaning tables before bulk loading.
Bulk Load stage	A passive stage whose role in a DataStage job is to take streams of tabular data and load them into tables of a target database.
control file	A file of commands for bulk loading one or more tables from a single link.
data file	An ASCII file of row/column data from an input link that is to be loaded.
local bulk loading	Bulk loading that occurs when both the DataStage job and Informix server are on the same physical machine.
remote bulk loading	Bulk loading that occurs when both the DataStage job and Informix server are on different physical machines.
Server	The Informix database to which you connect in order to use the Informix Load stage.

Installing Informix Load

For instructions and information supporting the installation, see *DataStage Plug-In Installation and Configuration Guide*.

After installing the server stage and creating a job, start the editor from the DataStage Designer. For more information about the editor, see DataStage documentation.

Local and Remote Processing

Informix Load supports both local bulk loading and remote bulk loading.

To achieve maximum bulk-loading performance, use the Informix Load stage locally on the same machine where the Informix database server is located. The default settings in the Informix Load stage assume the local bulk-loading processing.

In the case of the local processing, the only required link-level settings for the Informix Load stage are the following:

- Database Server Name
- Database Name
- Table Name (if the custom control file is not present)

See "Link Properties" on page 6 for details.

Additional link-level properties, however, may be used to customize or increase the scope of the Informix Load stage, for example, allowing for remote bulk loading.

Defining Data Types

The following sections describe DATETIME and DATE definitions.

DATETIME Considerations

In Informix, a valid DATETIME data type can be either *yyyy-mo-dd hh:mi:ss:ff*… or any contiguous subset of the previous string, for example, *mo-dd hh*. Any column that is defined to contain such a subset does not accept a full ANSI DATETIME.

Therefore, the DATETIME type generated within the data file corresponds to the DATETIME type defined for the actual column in a table. To be consistent with the

Informix DDL, you should define the following in the derivation for every DATETIME column (except the default):

Expression	Description
Hour to Fraction(5)	A partial ANSI DATETIME type, starting with hours and ending with fractions, with a precision of 5.
Hour to Second	A partial ANSI DATETIME type, starting with hours and ending with seconds.
Month to Minute	A partial ANSI DATETIME type, starting with months and ending with minutes.
Year to Fraction(3)	The default value. Full ANSI DATETIME, with a precision of 3.

DATE Considerations

In Informix, the interpretation of a DATE data type by a DBLOAD bulk loader depends on the settings of the DBDATE or GL_DATE environment variables. For example, if DBDATE is set to DMY4, then the *xx-yy-zzzz* string is interpreted as the *xx* day of the *yy* month for the *zzzz* year.

Informix Load cannot read the DBDATE or GL_DATE environment variables automatically for date formatting for the DBLOAD utility, since the DataStage job can be run on a different physical machine than the one where the Informix server is installed.

The default DBDATE Informix setting is MDY4. DBLOAD converts the date *xxxx-yy-zz* format only if the DBDATE setting is Y4MD; otherwise, an error occurs converting the string to a date.

Regardless of the DBDATE setting, the DBLOAD utility correctly interprets the following:

- Date separator as - or /
- Two-digit rather than four-digit year (MDY4 is equivalent to MDY2 for dates, for example, with the *xx-yy-zz* format)

Link Properties

Informix Load supports the following link properties that are visible from the DataStage Manager. The following table includes these column heads:

- **Property** is the program-visible name of the property.
- **Prompt** is the text that you see in the stage editor user interface.
- **Default** is the text used if you do not supply any value.
- **Help Text** is added to the table to describe the properties (this text is not visible using the Informix Load stage).

Prompt	Type	Default	Description
Table Name	String	None	(Required if Control File is not specified) Table name to load into.
Database Name	String	None	(Required) The name of the database on which the target table resides.
Server Name	String	None	(Required) The name of the Informix server on which the target table resides.
Run DBLOAD	String List	Yes	(Yes, No) Specifies whether to run a local DBLOAD utility automatically or not. If a DataStage Informix Load job is on a different physical machine than the Informix database, this setting should be set to No.
Hex Escapes	String List	Yes	(Yes, No) Recognize hexadecimal escapes in character fields.
Table Locking	String List	Yes	(Yes, No) Specifies whether to provide for regular table locking during bulk loading.
Exclusive locking	String List	Yes	(Yes, No) Specifies whether to provide for exclusive table locking during bulk loading.
Field Delimiter	String	None	Field delimiter character. To represent the delimiter itself, precede it by a backslash (\). To represent a backslash, double it. If you use a backslash in your data, whether NLS is off or on, you must populate Field Delimiter in order for a proper conversion to occur.

Prompt	Type	Default	Description
Before-link Routine	String	None	Any valid name of the external process or routine with arguments referring to link property values (for example, *xxx.bat %CTRFILE% %DATAFILE% %ERRLOG%*). The arguments within the percent (%) delimiters correspond to program-visible property names. This process executes before the link activity.
After-link Routine	String	None	Any valid name of the external process or routine with arguments, containing link property values (for example, *Postlink.bat %CTRFILE% %DATAFILE% %ERRLOGFILE%*). The arguments within the percent (%) delimiters correspond to program-visible property names. This process executes after completion of all link activity.
Control File	String	None	If this property contains a valid custom control file name, this file is used, and the Table Name property is ignored. Otherwise, a file *database_table.ctl* is generated (the file name combines the database name and table name property values).
Data File	String	None	If this property contains a valid file name, this file is generated. Otherwise, a file *database_table.dat* is generated (the file name combines the database name and table name property values).
Errorlog File	String	None	If this property contains a valid file name, this file is generated. Otherwise, a unique file *database_table.log* is generated (the file name combines the database name and table name property values).
Directory Name	String	None	Full directory path to place all generated files. If empty, the current directory is assumed.
Datetime Format	String	None	Informix DDL-based default DATETIME data type format for all the columns in the table. Individual columns can override this setting in their description attribute.

Prompt	Type	Default	Description
Date Format	Long	None	Valid date formats are: 1 – *mm-dd-yyyy* 2 – *dd-mm-yyyy* 3 – *yyyy-mm-dd* The DATE type should correspond to the type defined through the DBDATE or GL_DATE environment variables.
Maximum Bad Rows	Long	10	Maximum errant rows before aborting bulk loading.
Rows Batch Size	Long	100	Number of rows to process before commit.
Rows to Ignore	Long	0	Starting rows to ignore.

For more information about Informix and Bulk Load utilities, see *Informix Migration Guide*.